

## Research Article

# Markov Switching Model Analysis of Implied Volatility for Market Indexes with Applications to S&P 500 and DAX

Luca Di Persio<sup>1</sup> and Samuele Vettori<sup>2</sup>

<sup>1</sup>Department of Computer Science, University of Verona, Strada le Grazie 15, 37134 Verona, Italy

<sup>2</sup>Department of Mathematics, University of Trento, Via Sommarive 14, 38123 Trento, Italy

Correspondence should be addressed to Luca Di Persio; [dipersioluca@gmail.com](mailto:dipersioluca@gmail.com)

Received 31 May 2014; Revised 26 November 2014; Accepted 26 November 2014; Published 18 December 2014

Academic Editor: Niansheng Tang

Copyright © 2014 L. Di Persio and S. Vettori. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We adopt a *regime switching* approach to study concrete financial time series with particular emphasis on their volatility characteristics considered in a space-time setting. In particular the volatility parameter is treated as an unobserved state variable whose value in time is given as the outcome of an unobserved, discrete-time and discrete-state, stochastic process represented by a suitable Markov chain. We will take into account two different approaches for inference on Markov switching models, namely, the classical approach based on the maximum likelihood techniques and the Bayesian inference method realized through a Gibbs sampling procedure. Then the classical approach shall be tested on data taken from the *Standard & Poor's 500* and the *Deutsche Aktien Index* series of returns in different time periods. Computations are given for a four-state switching model and obtained numerical results are put beside by explanatory graphs which report the outcomes obtained exploiting both smoothing and filtering algorithms used in the estimation/calibration procedures we proposed to infer on the switching model parameters.

## 1. Introduction

Many financial time series are characterized by abrupt changes in their behaviour, a phenomena that can be implied by a number of both endogeneous and exogeneous facts, often far from being forecasted. Examples of such changing factors can be represented, for example, by large financial crises, government policy and political instabilities, natural disasters, and speculative initiatives.

Such phenomena have been frequently observed during last decade especially because of the worldwide financial crisis which originated during the first years of 2000 and is still running. Indeed such a crisis, often referred to as the *Global Financial Crisis*, has caused a big lack of liquidity for banks, both in USA, Europe, and many countries all over the world, resulting, for example, in a collapse of many financial institutions, a generalized downfall in stock markets, a decline consumer wealth, and an impressive growth of the European sovereign debt. The reasons behind these phenomena are rather complicated, particularly because of the high number of interconnected point of interests, each of which is

driven by specific influences often linked between each other. Nevertheless there are mathematical techniques which can be used to point out some general characteristics able to synthesize some relevant informations and to give an effective help in forecasting future behaviour of certain macroquantities of particular interest.

Although linear time series techniques, for example, the autoregressive (AR) model, the moving average (MA) model, and their combination (ARMA), have been successfully applied in a large number of financial applications, by their own nature they are unable to describe nonlinear dynamic patterns as in the case, for example, of asymmetry and volatility clustering. In order to overcome latter issue various approaches have been developed. Between them and with financial applications in mind, we recall the autoregressive conditional heteroskedasticity (ARCH) model of Engle, together with its generalised version (GARCH), and the *regime switching* (RS) model which involves multiple equations, each characterizing the behaviors of quantities of interest in different regimes, and a mechanism allowing the switch between them. Concerning the switching methods

that can be considered in the RS framework, we would like to cite the *threshold autoregressive* (TAR) model proposed by Tong in [1], in which regime switching is controlled by a fixed threshold, the *autoregressive conditional root* (ACR) model of Bec et al. (see [2]) where the regime switching between stationary and nonstationary state is controlled by a binary random variable, and its extension, namely, the *functional coefficient autoregressive conditional root* (FCACR) model, considered by Zhou and Chen in [3]. In particular in this work we aim at using the RS approach to model aforementioned types of unexpected changes by their dependence on an unobserved variable typically defined as the *regime* or *state*. A customary way to formalize such an approach is given through the following *state-space* representation:

$$y_t = f(S_t, \theta\psi), \quad t = 1, \dots, n, \quad n \in \mathbb{N}^+, \quad S_t \in \Omega, \quad (1)$$

where  $\theta$  is the vector of problem parameters,  $\psi$  is the information set, the *state set*  $\Omega$ , with  $|\Omega| = M \in \mathbb{N}^+$ , is the (finite) set of possible values which can be taken by the state process  $S$  at time  $t$ , that is,  $S_t \in \Omega$ , and  $f$  is a suitable function determining the value of the dependent variable  $y$  at any given time  $t \in \{1, \dots, n\}$ .

The simplest type of structure considers two regimes; that is,  $\Omega = \{1, 2\}$  and at most one switch in the time series: in other words, the first  $m$  (unknown) observations relate regime 1, while the remaining  $n - m$  data concern regime 2. Such an approach can be generalized allowing the system to switch back and forth between the two regimes, with a certain probability. The latter is the case considered by Quandt in his paper published in 1962, where it is assumed to be theoretically possible for the system to switch between regimes every time that a new observation is generated. Note that previous hypothesis is not realistic in an economic context, since it contradicts the volatility clustering property, typical of financial time series.

The best way to represent the volatility clusters phenomenon consists in assuming that the state variable follows a Markov chain and claiming that the probability of having a switch in the next time is much lower than the probability of remaining in the same economic regime. The Markovian switching mechanism was first considered by Goldfeld and Quandt in [4] and then extended by Hamilton to the case of structural changes in the parameters of an autoregressive process (see [5]). When the unobserved state variable that controls the switching mechanism follows a first-order Markov chain, the RS model is called Markov Switching Model (MSM). In particular the Markovian property of such a model implies that, given  $t \in \{2, \dots, n\}$ , the value  $S_t$  of the state variable depends only on  $S_{t-1}$ , a property that turns out to be useful to obtain a good representation of financial data, where abrupt changes occur occasionally.

After Hamilton's papers, Markov switching models have been widely applied, together with a number of alternative versions, to analyze different type of both economic and financial time series, for example, stock options behaviors, equity markets trends, and interest rates series. In what

follows we shall often refer to the following, actually rather general, form for an MSM:

$$y_t = f(S_t, \theta)$$

$$S_t \in \{1, 2, \dots, M\} \quad \text{with probabilities}$$

$$p_{ij} := \mathbb{P}(S_t = j \mid S_{t-1} = i), \quad i, j = 1, 2, \dots, M, \quad t = 2, \dots, n, \quad (2)$$

where the terms  $p_{ij}$  are nothing but the transition probabilities, from state  $i$  at time  $t - 1$  to state  $j$  at time  $t$ , which determine the stochastic dynamic of the process  $S = \{S_t\}_{t=1, \dots, n}$ .

In Section 2 we recall the classical approach to MSM following [5] and with respect to both serially uncorrelated and correlated data. Then, in Section 3, we first introduce basic facts related to the Bayesian inference; then we recall the Gibbs sampling technique and related Monte Carlo approximation method which is later used to infer on the MSM parameters. Section 4 is devoted to a *goodness of fit* (of obtained estimates for parameters of interest) analysis, while in Section 5 we describe our forecasting MSM-based procedure which is then used in Section 6 to analyse both the *Standard & Poor's 500* and the *Deutsche Aktien* indexes.

## 2. The Classical Approach

In this section we shall use the *classical approach* (see, for example, [5]) to develop procedures which allow us to make inference on unobserved variables and parameters characterizing MSM. The main idea behind such an approach is split in two steps: first we estimate the model's unknown parameters by a maximum likelihood method; secondly we infer the unobserved switching variable values conditional on the parameter estimates. Along latter lines, we shall analyze two different MSM settings, namely, the case in which data are *serially uncorrelated*, and the case when they are *autocorrelated*.

**2.1. Serially Uncorrelated Data.** Let us suppose that  $S = \{S_t\}_{t \in \{1, \dots, T\}}$ ,  $T \in \mathbb{N}^+$ , is a discrete time stochastic process represented by a first-order Markov chain taking value in some set  $\Omega = \{1, \dots, M\}$ , with transition probabilities:

$$\begin{aligned} p_{ij} &:= \mathbb{P}(S_{t+1} = j \mid S_t = i, S_{t-1} = i_{t-1}, \dots, S_0 = i_0) \\ &= \mathbb{P}(S_{t+1} = j \mid S_t = i), \quad \forall t \in \{2, \dots, T\}. \end{aligned} \quad (3)$$

Then the state-space model we want to study is given by

$$\begin{aligned} y_t &= \mu_{S_t} + \epsilon_t, \quad t = 1, 2, \dots, T, \\ \{\epsilon_t\}_{t \in \{1, \dots, T\}} &\text{ are i.i.d. } \mathcal{N}(0, \sigma_{S_t}^2), \\ \mu_{S_t} &= \mu_1 S_{1,t} + \dots + \mu_M S_{M,t}, \\ \sigma_{S_t} &= \sigma_1 S_{1,t} + \dots + \sigma_M S_{M,t}, \\ S_t &\in \{1, \dots, M\} \quad \text{with probabilities } p_{ij}, \end{aligned} \quad (4)$$

where the variables  $S_{k,t}$ ,  $(k, t) \in \{1, \dots, M\} \times \{1, \dots, T\}$  are introduced in order to have a slightly compact equation for  $\mu_{S_t}$  and  $\sigma_{S_t}$ : in particular,  $S_{k,t} = 1$  if  $S_t = k$ , otherwise  $S_{k,t} = 0$ , which implies that under regime  $k$ , for  $k \in \{1, \dots, M\}$ , parameters of mean and variance are given by  $\mu_k$  and  $\sigma_k$ .

Let us underline that in (4) the  $y_t$  are our observed data, for example, historical returns of a stock or some index time series, and we suppose that they are *locally* distributed as Gaussian random variable in the sense that occasionally jumps could occur for both the mean  $\mu_{S_t}$  and the variance  $\sigma_{S_t}^2$ . In particular, we assume that  $\sigma_1 < \sigma_2 < \dots < \sigma_M$  and we want to estimate these  $M$  unobserved values for standard deviation, as well as the  $M$  values for the switching mean. Note that we could also take  $\mu$  as a constant, obtaining the so called *switching variance problem* or  $\sigma$  as a constant, having a *switching mean problem*. The first one is in general more interesting in the analysis of financial time series, since variance is usually interpreted as an indicator for the market's volatility.

Given the model described by (4), the conditional density of  $y_t$  given  $S_t$  is Gaussian with mean  $\mu_{S_t}$  and standard deviation  $\sigma_{S_t}$ ; namely, its related probability density function reads as follow:

$$f_{y_t|S_t}(x) = \frac{1}{\sqrt{2\pi\sigma_{S_t}^2}} e^{-(x-\mu_{S_t})^2/2\sigma_{S_t}^2}, \quad x \in \mathbb{R}, \quad (5)$$

and we are left with the problem of estimating both the expectations  $\mu_1, \dots, \mu_M$  and the standard deviations  $\sigma_1, \dots, \sigma_M$  parameters; a task that is standard to solve by maximizing the associated log-likelihood function  $\ln \mathcal{L}$  defined as follows:

$$\ln \mathcal{L} := \sum_{t=1}^T \ln(f_{y_t|S_t}). \quad (6)$$

A different and more realistic scenario is the one characterized by unobserved values for  $S_t$ . In such a case, it could be possible to consider the MSM-inference problem as a two-step procedure consisting in

- (i) estimating the parameters of the model by maximizing the log-likelihood function,
- (ii) making inferences on the state variable  $S_t$ ,  $t = 1, \dots, T$  conditional on the estimates obtained at previous point.

Depending on the amount of information we can use inferring on  $S_t$ , we have

- (i) *filtered probabilities* that refer to inferences about  $S_t$  conditional on information up to time  $t$ , namely, with respect to  $\psi_t$ ,
- (ii) *smoothed probabilities* that refer to inferences about  $S_t$  conditional to the whole sample (*history*),  $\psi_T$ .

In what follows we describe a step-by-step algorithm which allows us to resolve the filtering problem for a sample of serially uncorrelated data. In particular we slightly generalize the approach given in [6], assuming that the state variable  $S_t$

belongs to a 4-state space set  $\Omega := \{1, 2, 3, 4\}$ , at every time  $t = 1, \dots, T$ . Despite 2-state (expansion/contraction) and 3-state (low/medium/high volatility regime) models being the usual choices, we decided to consider 4-state MSM in order to refine the analysis with respect to volatility levels around the mean. A finer analysis can be also performed, even if one has to take into account the related nonlinear computational growth. We first define the log-likelihood function at time  $t$  as

$$l(\theta, \psi_t) = l(\mu_1, \dots, \mu_M, \sigma_1, \dots, \sigma_M, \psi_t), \quad (7)$$

where  $\theta := (\mu_1, \dots, \mu_M, \sigma_1, \dots, \sigma_M)$  is the vector of parameters that we want to estimate. Let us note that the  $\theta$  is updated at every iteration, since we maximize with respect to the function  $l(\theta, \psi_t)$  at every stage of our step-by-step procedure. In particular the calibrating procedure reads as follows.

*Inputs.*

- (i) Put  $l(\theta) := l(\theta, \psi_0) = 0$ .
- (ii) Compute the *transition probabilities* of the homogeneous Markov chain underlying the state variable; that is,

$$p_{ij} = \mathbb{P}(S_t = j \mid S_{t-1} = i), \quad i, j \in \Omega. \quad (8)$$

Since in the applications we can only count on return time series, we first have to calibrate with respect to the transition probabilities  $p_{ij}$ :

- (1) choose 4 values  $\sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2$  (e.g.,  $\sigma_1 = 0.1$ ,  $\sigma_2 = 0.2$ ,  $\sigma_3 = 0.3$ , and  $\sigma_4 = 0.4$ ) and a positive, arbitrary small, constant  $\delta > 0$ , for example,  $\delta = 0.01$ ,
- (2) compute for every  $j = 1, 2, 3, 4$  and for every  $t = 1, \dots, T$  the values

$$b_j := \Phi(y_t + \delta) - \Phi(y_t - \delta) = \int_{y_t - \delta}^{y_t + \delta} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left[-\frac{(x - \mu_j)^2}{2\sigma_j^2}\right], \quad (9)$$

- (3) simulate a value in  $\{1, 2, 3, 4\}$  for  $S_t$  at each time from the discrete probability vector

$$\left( \frac{b_1}{\sum_{j=1}^4 b_j}, \frac{b_2}{\sum_{j=1}^4 b_j}, \frac{b_3}{\sum_{j=1}^4 b_j}, \frac{b_4}{\sum_{j=1}^4 b_j} \right), \quad (10)$$

- (4) set the transition probabilities  $p_{ij}$  just by counting the number of transition from state  $i$  to state  $j$ , for  $i, j = 1, \dots, 4$ , in order to obtain the following *transition matrix*:

$$P^* = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix}. \quad (11)$$

(iii) Compute the *steady-state probabilities*

$$\pi(0) = (\mathbb{P}(S_0 = 1 | \psi_0), \dots, \mathbb{P}(S_0 = 4 | \psi_0)). \quad (12)$$

Let us note that, by definition, if  $\pi(t)$  is a  $4 \times 1$  vector of steady-state probabilities, then  $\pi(t+1) = \pi(t)$  for every  $t = 1, \dots, T$ ; moreover  $\pi(t+1) = P^* \pi(t)$ , and (see for example, [6, pag. 71]) we also have that  $\pi(t) = (A^T A)^{-1} A^T \begin{bmatrix} 0_4 \\ 1 \end{bmatrix}$ , where  $0_4$  is a  $4 \times 1$  matrix of zeros and  $A = \begin{pmatrix} \text{Id}_4 - P^* \\ 1_4 \end{pmatrix}$ ,  $\text{Id}_4$  is the four dimensional identity matrix, while  $1_4 = (1, 1, 1, 1)$ , that is, the vector of steady-state probabilities is the last column of the matrix  $(A^T A)^{-1} A^T$ .

Next we perform the following steps, for  $t = 1, \dots, T$ .

*Step 1.* The probability of  $S_t$  conditional to information set at time  $t-1$  is given by

$$\mathbb{P}(S_t = j | \psi_{t-1}) = \sum_{i=1}^4 p_{ij} \mathbb{P}(S_{t-1} = i | \psi_{t-1}), \quad j = 1, \dots, 4. \quad (13)$$

*Step 2.* Compute the joint density of  $y_t$  and  $S_t$  conditional to the information set  $\psi_{t-1}$

$$\begin{aligned} f(y_t, S_t = j | \psi_{t-1}) &= f(y_t | S_t = j, \psi_{t-1}) \\ &\times \mathbb{P}(S_t = j | \psi_{t-1}), \quad j = 1, \dots, 4. \end{aligned} \quad (14)$$

The marginal density of  $y_t$  is given by the sum of the joint density over all values of  $S_t$

$$f(y_t | \psi_{t-1}) = \sum_{i=1}^4 f(y_t | S_t = i, \psi_{t-1}) \mathbb{P}(S_t = i | \psi_{t-1}). \quad (15)$$

*Step 3.* Update the log-likelihood function at time  $t$  in the following way:

$$l(\theta, \psi_t) = l(\theta, \psi_{t-1}) + \ln(f(y_t | \psi_{t-1})), \quad (16)$$

and maximize  $l(\theta, \psi_t)$  with respect to  $\theta = (\mu_1, \dots, \mu_4, \sigma_1, \dots, \sigma_4)$ , under the condition  $\sigma_1 < \sigma_2 < \sigma_3 < \sigma_4$ , to find the maximum likelihood estimator  $\hat{\theta}$  for the next time period.

*Step 4.* Once  $y_t$  is observed at the end of the  $t$ th iteration, we can *update* the probability term:

$$\begin{aligned} \mathbb{P}(S_t = j | \psi_t) &= \frac{f(y_t, S_t = j, \psi_{t-1}) \mathbb{P}(S_t = j | \psi_{t-1})}{\sum_{i=1}^4 f(y_t | S_t = i, \psi_{t-1}) \mathbb{P}(S_t = i | \psi_{t-1})}, \\ &\quad j = 1, \dots, 4, \end{aligned} \quad (17)$$

where both  $\psi_t = \{\psi_{t-1}, y_t\}$  and  $f(y_t | S_t = i, \psi_{t-1})$  are computed with respect to the estimator  $\hat{\theta} = (\hat{\mu}_1, \dots, \hat{\mu}_4, \hat{\sigma}_1, \dots, \hat{\sigma}_4)$ .

*2.2. Serially Correlated Data.* In some cases it is possible to argue and mathematically test by, for example, the Durbin-Watson statistics or Breusch-Godfrey test, for the presence of a serial correlation (or *autocorrelation*) between data belonging to a certain time series of interest. Such a characteristic is often analyzed in signal processing scenario, but examples can be also found in economic, meteorological, or sociological data sets, especially in connection with autocorrelation of errors in related forecasting procedure. In particular if we suppose that the observed variable  $y_t$  linearly depends on its previous value, then we obtain a *first-order autoregressive pattern* and the following state-space model applies:

$$y_t - \mu_{S_t} = \phi(y_{t-1} - \mu_{S_{t-1}}) + \epsilon_t, \quad t = 1, 2, \dots, T,$$

$$\{\epsilon_t\}_{t \in \{1, \dots, T\}} \text{ are i.i.d. } \mathcal{N}(0, \sigma_{S_t}^2),$$

$$\mu_{S_t} = \mu_1 S_{1,t} + \dots + \mu_M S_{M,t}, \quad (18)$$

$$\sigma_{S_t} = \sigma_1 S_{1,t} + \dots + \sigma_M S_{M,t},$$

$$S_t \in \{1, \dots, M\} \text{ with probabilities } p_{ij},$$

where  $p_{ij} := \mathbb{P}(S_t = j | S_{t-1} = i)$ ,  $i, j = 1, \dots, M$  and  $S_{k,t}$ , and  $(k, t) \in \{1, \dots, M\} \times \{1, \dots, T\}$  are the same variables introduced in the previous section; that is,  $S_{k,t} = 1$  if  $S_t = k$ , otherwise  $S_{k,t} = 0$ .

In this situation, if the state  $S_t$  is known for every  $t = 1, \dots, T$ , we need  $S_t$  and  $S_{t-1}$  to compute the density of  $y_t$  conditional to past information  $\psi_{t-1}$ , indeed we have

$$\ln \mathcal{L} := \sum_{t=1}^T \ln(f_{y_t | \psi_{t-1}, S_t, S_{t-1}}), \quad (19)$$

where

$$f_{y_t | \psi_{t-1}, S_t, S_{t-1}}(x) = \frac{1}{\sqrt{2\pi\sigma_{S_t}^2}} e^{-(x - \mu_{S_t} - \phi(y_{t-1} - \mu_{S_{t-1}}))^2 / 2\sigma_{S_t}^2}, \quad x \in \mathbb{R}. \quad (20)$$

If  $S_t$  are unobserved (and as before we assume that the state variable can take the four values  $\{1, 2, 3, 4\}$ ), we apply the following algorithm in order to resolve the filtering problem for a sample of serially correlated data.

*Inputs.*

(i) Put  $l(\theta) := l(\theta, \psi_0) = 0$ .

(ii) Compute the transition probabilities

$$p_{ij} = \mathbb{P}(S_t = j | S_{t-1} = i), \quad i, j = 1, 2, 3, 4. \quad (21)$$

We apply the same trick as before, but firstly we have to estimate the parameter  $\phi$ : in order to obtain this value we can use the least square methods (see for example, [7]) that is,

$$\hat{\phi} := \frac{\sum_{t=1}^T y_t y_{t+1}}{\sum_{t=1}^T y_t^2}. \quad (22)$$



Then we compute  $z_t := y_t - \phi y_{t-1}$  for every  $t = 1, \dots, T$  and consider the values  $b_j := \Phi(z_t + \delta) - \Phi(z_t - \delta)$  (we apply the Normal distribution function to  $z_t + \delta$  instead of  $y_t + \delta$ , as done before).

(iii) Compute the steady-state probabilities

$$\pi(0) = (\mathbb{P}(S_0 = 1 | \psi_0), \dots, \mathbb{P}(S_0 = 4 | \psi_0)), \quad (23)$$

taking the last column of the matrix  $(A^T A)^{-1} A^T$  (see procedure in Section 2.1 for details).

Next perform the following steps for  $t = 1, \dots, T$ .

*Step 1.* Compute the probabilities of  $S_t$  conditional to information set at time  $t - 1$ , for  $j = 1, \dots, 4$

$$\begin{aligned} \mathbb{P}(S_t = j | \psi_{t-1}) &= \sum_{i=1}^4 \mathbb{P}(S_t = j, S_{t-1} = i | \psi_{t-1}) \\ &= \sum_{i=1}^4 p_{ij} \mathbb{P}(S_{t-1} = i | \psi_{t-1}). \end{aligned} \quad (24)$$

*Step 2.* Compute the joint density of  $y_t$ ,  $S_t$ , and  $S_{t-1}$  given  $\psi_{t-1}$ :

$$\begin{aligned} f(y_t, S_t, S_{t-1} | \psi_{t-1}) &= f(y_t | S_t = j, S_{t-1} = i, \psi_{t-1}) \\ &\quad \times \mathbb{P}(S_t = j, S_{t-1} = i | \psi_{t-1}), \end{aligned} \quad (25)$$

where  $f(y_t | S_t = j, S_{t-1} = i, \psi_{t-1})$  is given by (20) and  $\mathbb{P}(S_t = j, S_{t-1} = i | \psi_{t-1})$  is computed in Step 1. The marginal density of  $y_t$  conditional on  $\psi_{t-1}$  is obtained by summing over all values of  $S_t$  and  $S_{t-1}$ :

$$\begin{aligned} f(y_t | \psi_{t-1}) &= \sum_{j=1}^4 \sum_{i=1}^4 f(y_t | S_t = j, S_{t-1} = i, \psi_{t-1}) \\ &\quad \times \mathbb{P}(S_t = j, S_{t-1} = i | \psi_{t-1}). \end{aligned} \quad (26)$$

*Step 3.* The log-likelihood function at time  $t$  is again

$$l(\theta, \psi_t) = l(\theta, \psi_{t-1}) + \ln(f(y_t | \psi_{t-1})), \quad (27)$$

and it can be maximized with respect to  $\theta = (\mu_1, \dots, \mu_4, \sigma_1, \dots, \sigma_4)$ , under condition  $\sigma_1 < \sigma_2 < \sigma_3 < \sigma_4$ , giving the maximum likelihood estimator  $\hat{\theta}$  for the next time period.

*Step 4.* Update the joint probabilities of  $S_t$  and  $S_{t-1}$  conditional to the new information set  $\psi_t$ , using the estimator  $\hat{\theta}$  computed in Step 3 by maximizing the log-likelihood function  $l(\theta, \psi_t)$ :

$$\begin{aligned} \mathbb{P}(S_t = j, S_{t-1} = i | \psi_t) &= \frac{f(y_t, S_t = j, S_{t-1} = i | \psi_{t-1})}{f(y_t | \psi_{t-1})}, \quad i, j = 1, \dots, 4. \end{aligned} \quad (28)$$

Then compute the updated probabilities of  $S_t$  given  $\psi_t$  by summing the joint probabilities over  $S_{t-1}$  as follows:

$$\mathbb{P}(S_t = j | \psi_t) = \sum_{i=1}^4 \mathbb{P}(S_t = j, S_{t-1} = i | \psi_t), \quad \forall j = 1, \dots, 4. \quad (29)$$

*The Smoothing Algorithm.* Once we have run this procedure, we are provided with the *filtered probabilities*, that is, the values  $\mathbb{P}(S_t = j | \psi_t)$  for  $j = 1, \dots, 4$  and for each  $t = 1, \dots, T$  (in addition to the estimator  $\hat{\theta}$ ).

Sometimes it is required to estimate probabilities of  $S_t$  given the whole sample information; that is,

$$\mathbb{P}(S_t = j | \psi_T) = \mathbb{P}(S_t = j | y_1, \dots, y_T), \quad \forall j = 1, \dots, 4, \quad (30)$$

which are called *smoothed probabilities*. We are going to show how these new probabilities can be computed from previous procedure (the same algorithm, although with some obvious changes, can be still used starting from procedure in Section 2.1).

Since the last iteration of the algorithm gives us the probabilities  $\mathbb{P}(S_T = j | \psi_T)$  for  $j = 1, \dots, 4$ , we can start from these values and use the following procedure by doing the two steps for every  $t = T - 1, T - 2, \dots, 2, 1$ .

*Step 1.* For  $i, j = 1, \dots, 4$ , compute

$$\begin{aligned} \mathbb{P}(S_t = i, S_{t+1} = j | \psi_T) &= \mathbb{P}(S_{t+1} = j | \psi_T) \mathbb{P}(S_t = i | S_{t+1} = j, \psi_T) \\ &= \mathbb{P}(S_{t+1} = j | \psi_T) \mathbb{P}(S_t = i | S_{t+1} = j, \psi_t) \quad (*) \\ &= \mathbb{P}(S_{t+1} = j | \psi_T) \frac{\mathbb{P}(S_t = i, S_{t+1} = j | \psi_t)}{\mathbb{P}(S_{t+1} = j | \psi_t)}. \end{aligned}$$

*Remark 1.* Note that equality (\*), that is,

$$\mathbb{P}(S_t = i | S_{t+1} = j, \psi_T) = \mathbb{P}(S_t = i | S_{t+1} = j, \psi_t), \quad (31)$$

holds only under a particular condition; namely,

$$f(h_{t+1,T} | S_{t+1} = j, S_t = i, \psi_t) = f(h_{t+1,T} | S_{t+1} = j, \psi_t), \quad (32)$$

where  $h_{t+1,T} := (y_{t+1}, \dots, y_T)'$  (see [6] for the proof). Equation (32) suggests that if  $S_{t+1}$  were known, then  $y_{t+1}$  would contain no information about  $S_t$  beyond that contained in  $S_{t+1}$  and  $\psi_t$  and does not hold for every state-space model with regime switching (see, for example, [6, Ch. 5]) in which case the smoothing algorithm involves an approximation.

*Step 2.* For  $i = 1, \dots, 4$  compute

$$\mathbb{P}(S_t = i | \psi_T) = \sum_{j=1}^4 \mathbb{P}(S_t = i, S_{t+1} = j | \psi_T). \quad (33)$$

### 3. The Gibbs Sampling Approach

*3.1. An Introduction to Bayesian Inference.* Under the general title *Bayesian inference* we can collect a large number of different concrete procedures; nevertheless they are all based on smart use of the Bayes' rule which is used to update the probability estimate for a hypothesis as additional evidence is learned (see, for example, [8, 9]). In particular, within

the Bayesian framework, the parameters, for example, let us collect them in a vector called  $\theta$ , which characterize a certain statistic model and are treated as random variables with their own probability distributions; let us say  $f(\theta)$ , which plays the role of a *prior distribution* since it is defined before taking into account the sample data  $\tilde{y}$ . Therefore, exploiting the Bayes' theorem and denoting by  $f(\tilde{y} | \theta)$  the likelihood of  $\tilde{y}$  of the interested statistic model, we have that

$$f(\theta | \tilde{y}) = \frac{f(\tilde{y} | \theta) f(\theta)}{f(\tilde{y})}, \quad (34)$$

where  $f(\theta | \tilde{y})$  is the *joint posterior distribution* of the parameters. The denominator  $f(\tilde{y})$  defines the marginal likelihood of  $\tilde{y}$  and can be taken as a constant, obtaining the proportion

$$f(\theta | \tilde{y}) \propto f(\tilde{y} | \theta) f(\theta). \quad (35)$$

It is straightforward to note that the most critical part of the Bayesian inference procedure relies in the choice of a suitable prior distribution, since it has to agree with parameters constraints. An effective answer to latter issue is given by the so called *conjugate prior distribution*, namely the distribution obtained when the conjugate prior is combined with the likelihood function. Let us note that the posterior distribution  $f(\theta | \tilde{y})$  is in the same family as the prior distribution.

As an example, if the likelihood function is Gaussian, it can be shown that the conjugate prior for the mean  $\mu$  is the Gaussian distribution, whereas the conjugate prior for the variance is the inverted Gamma distribution (see, for example, [9, 10]).

**3.2. Gibbs Sampling.** A general problem in Statistics concerns the question of how a sequence of observations which cannot be directly sampled, can be simulated, for example, by mean of some multivariate probability distribution, with a prefixed precision degree of accuracy. Such kind of problems can be successfully attacked by Monte Carlo Markov Chain (MCMC) simulation methods, see, for example, [11–13], and in particular using the so called *Gibbs Sampling* technique which allows to approximate joint and marginal distributions by sampling from conditional distributions, see, for example, [14–16].

Let us suppose that we have the joint density of  $k$  random variables, for example,  $f = f(z_1, z_2, \dots, z_k)$ , fix  $t \in \{1, \dots, k\}$  and that we are interested in obtaining characteristics of the  $z_t$ -marginal, namely

$$f(z_t) = \int \cdots \int f(z_1, z_2, \dots, z_k) dz_1 \cdots dz_{t-1} dz_{t+1} \cdots dz_k, \quad (36)$$

such as the related mean and/or variance. In those cases when the joint density is not given, or the above integral turns out to be difficult to treat, for example, an explicit solution does not exist, but we know the complete set of conditional densities, denoted by  $f(z_t | z_{j \neq t})$ ,  $t = 1, 2, \dots, k$ , with  $z_{j \neq t} = \{z_1, \dots, z_{t-1}, z_{t+1}, \dots, z_k\}$ , then the Gibbs Sampling method allows us to generate a sample  $z_1^j, z_2^j, \dots, z_k^j$  from the

joint density  $f(z_1, z_2, \dots, z_k)$  without requiring that we know either the joint density or the marginal densities. With the following procedure we recall the basic ideas on which the Gibbs Sampling approach is based given an arbitrary starting set of values  $(z_1^0, \dots, z_k^0)$ .

*Step 1.* Draw  $z_1^1$  from  $f(z_1 | z_2^0, \dots, z_k^0)$ .

*Step 2.* Draw  $z_2^1$  from  $f(z_2 | z_1^1, z_3^0, \dots, z_k^0)$ .

*Step 3.* Draw  $z_3^1$  from  $f(z_3 | z_1^1, z_2^1, z_4^0, \dots, z_k^0)$ .

$\vdots$

*Step k.* Finally draw  $z_k^1$  from  $f(z_k | z_1^1, \dots, z_{k-1}^1)$  to complete the first iteration.

The steps from 1 through  $k$  can be iterated  $J$  times to get  $(z_1^j, z_2^j, \dots, z_k^j)$ ,  $j = 1, 2, \dots, J$ .

In [17] S. Geman and D. Geman showed that both the joint and marginal distributions of generated  $(z_1^j, z_2^j, \dots, z_k^j)$ , converge, at an exponential rate, to the joint and marginal distributions of  $z_1, z_2, \dots, z_k$ , as  $J \rightarrow \infty$ . Thus the joint and marginal distributions of  $z_1, z_2, \dots, z_k$  can be approximated by the empirical distributions of  $M$  simulated values  $(z_1^j, z_2^j, \dots, z_k^j)$ ,  $j = L + 1, \dots, L + M$ , where  $L$  is large enough to assure the convergence of the Gibbs sampler. Moreover  $M$  can be chosen to reach the required precision with respect to the empirical distribution of interest.

In the MSM framework we do not have conditional distributions  $f(z_t | z_{j \neq t})$ ,  $t = 1, 2, \dots, k$ , and we are left with the problem of estimate parameters  $z_i$ ,  $i = 1, \dots, k$ . Latter problem can be solved exploiting Bayesian inference results, as we shall state in the next section.

**3.3. Gibbs Sampling for Markov Switching Models.** A major problem when dealing with inferences from Markov switching models relies in the fact that some parameters of the model are dependent on an unobserved variable, let us say  $S_t$ . We saw that in the classical framework, inference on Markov switching models consists first in estimating the model's unknown parameters via maximum likelihood, then inference on the unobserved Markov switching variable  $\tilde{S}_T = (S_1, S_2, \dots, S_T)$ , conditional on the parameter estimates, has to be performed.

In the Bayesian analysis, both the parameters of the model and the switching variables  $S_t$ ,  $t = 1, \dots, T$  are treated as random variables. Thus, inference on  $\tilde{S}_T$  is based on a joint distribution, no more on a conditional one. By employing Gibbs sampling techniques, Albert and Chib (see [14]) provided an easy to implement algorithm for the Bayesian analysis of Markov switching models. In particular in their work the parameters of the model and  $S_t$ ,  $t = 1, \dots, T$  are treated as missing data, and they are generated from appropriate conditional distributions using Gibbs sampling method. As an example, let us consider the following simple model with two-state Markov switching mean and variance:

$$y_t = \mu_{S_t} + \epsilon_t, \quad t = 1, 2, \dots, T,$$

$\{\epsilon_t\}_{t \in \{1, \dots, T\}}$  are i.i.d.  $\mathcal{N}(0, \sigma_{\epsilon_t}^2)$ ,

$$\mu_{S_t} = \mu_0 + \mu_1 S_t,$$

$$\sigma_{S_t} = \sigma_0^2 (1 - S_t) + \sigma_1^2 S_t = \sigma_0^2 (1 + h_1 S_t), \quad h_1 > 0, \quad (37)$$

where  $S_t \in \{0, 1\}$  with transition probabilities  $p := \mathbb{P}(S_t = 0 \mid S_{t-1} = 0)$ ,  $q := \mathbb{P}(S_t = 1 \mid S_{t-1} = 1)$ . The Bayesian method consider both  $S_t$ ,  $t = 1, \dots, T$  and the model's unknown parameters  $\mu_0, \mu_1, \sigma_0, \sigma_1, p$  and  $q$ , as random variables. In order to make inference about these  $T + 6$  variables, we need to derive the joint posterior density  $f(\tilde{S}_T, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q \mid \psi_T)$ , where  $\psi_T := (y_1, y_2, \dots, y_T)$  and  $\tilde{S}_T = (S_1, S_2, \dots, S_T)$ . Namely the realization of the Gibbs sampling relies on the derivation of the distributions of each of the above  $T + 6$  variables, conditional on all the other variables. Therefore we can approximate the joint posterior density written above by running the following procedure  $L + M$  times, where  $L$  is an integer large enough to guarantee the desired convergence. Hence we have the following scheme.

*Step 1.* We can derive the distribution of  $S_t$ ,  $t = 1, \dots, T$  conditional on the other parameters in two different ways.

- (1) *Single-move gibbs sampling:* generate each  $S_t$  from  $f(S_t \mid \tilde{S}_{\neq t}, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q, \psi_T)$ ,  $t = 1, \dots, T$ , where  $\tilde{S}_{\neq t} = (S_1, \dots, S_{t-1}, S_{t+1}, \dots, S_T)$ .
- (2) *Multi-move gibbs sampling:* generate the whole block  $\tilde{S}_T$  from  $f(\tilde{S}_T \mid \mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q, \psi_T)$ .

*Step 2.* Generate the transition probabilities  $p$  and  $q$  from  $f(p, q \mid \tilde{S}_T)$ . Note that this distribution is conditioned only on  $\tilde{S}_T$  because we assume that  $p$  and  $q$  are independent of both the other parameters of the model and the data,  $\psi_T$ .

If we choose the *Beta distribution* as prior distribution for both  $p$  and  $q$ , we have that posterior distribution  $f(p, q \mid \tilde{S}_T) = f(p, q) L(p, q \mid \tilde{S}_T)$  is again a Beta distribution. So, Beta distribution is a conjugate prior for the likelihood of transition probabilities.

*Step 3.* Generate  $\mu_0$  and  $\mu_1$  from  $f(\mu_0, \mu_1 \mid \tilde{S}_T, \sigma_0^2, \sigma_1^2, p, q, \psi_T)$ . In this case the conjugate prior is the *Normal distribution*.

*Step 4.* Generate  $\sigma_0^2$  and  $\sigma_1^2$  from  $f(\sigma_0^2, \sigma_1^2 \mid \tilde{S}_T, \mu_0, \mu_1, p, q, \psi_T)$ . From definition of the model we have that  $\sigma_1^2 = \sigma_0^2(1 + h_1)$ : we can first generate  $\sigma_0^2$  conditional on  $h_1$ , and then generate  $h_1 = 1 + h_1$  conditional on  $\sigma_0^2$ . We use in both cases the *Inverted Gamma distribution* as conjugate prior for the parameters.

For a more detailed description of these steps (see [6, pp. 211–218]). Here we examine only the so called *Multi-move Gibbs sampling*, originally motivated by Carter and Kohn (see [15]) in the context of state space models and then implemented in [6] for a MSM. For the sake of simplicity,

let us suppress the conditioning on model's parameters and denote

$$f(\tilde{S}_T \mid \psi_T) := f(\tilde{S}_T \mid \mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q, \psi_T). \quad (38)$$

Using the Markov property of  $\{S_t\}_{t \in \{1, \dots, T\}}$  it can be seen that

$$f(\tilde{S}_T \mid \psi_T) = f(S_T \mid \psi_T) \prod_{t=1}^{T-1} f(S_t \mid S_{t+1}, \psi_t), \quad (39)$$

where  $f(S_T \mid \psi_T) = \mathbb{P}(S_T \mid \psi_T)$  is provided by the last iteration of filtering algorithm (see Sections 2.1 and 2.2). Note that (39) suggests that we can first generate  $S_T$  conditional on  $\psi_T$  and then, for  $t = T - 1, T - 2, \dots, 1$ , we can generate  $S_t$  conditional on  $\psi_t$  and  $S_{t+1}$ , namely we can run the following steps.

*Step 1.* Run the basic filter procedure to get  $f(S_t \mid \psi_t)$ ,  $t = 1, 2, \dots, T$  and save them; the last iteration of the filter gives us the probability distribution  $f(S_T \mid \psi_T)$ , from which  $S_T$  is generated.

*Step 2.* Note that

$$\begin{aligned} f(S_t \mid S_{t+1}, \psi_t) &= \frac{f(S_t, S_{t+1} \mid \psi_t)}{f(S_{t+1} \mid \psi_t)} = \frac{f(S_{t+1} \mid S_t, \psi_t) f(S_t \mid \psi_t)}{f(S_{t+1} \mid \psi_t)} \\ &= \frac{f(S_{t+1} \mid S_t) f(S_t \mid \psi_t)}{f(S_{t+1} \mid \psi_t)} \propto f(S_{t+1} \mid S_t) f(S_t \mid \psi_t), \end{aligned} \quad (40)$$

where  $f(S_{t+1} \mid S_t)$  is the transition probability and  $f(S_t \mid \psi_t)$  has been saved from Step 1. So we can generate  $S_t$  in the following way, first calculate

$$\begin{aligned} \mathbb{P}(S_t = 1 \mid S_{t+1}, \psi_t) &= \frac{f(S_{t+1} \mid S_t = 1) f(S_t = 1 \mid \psi_t)}{\sum_{j=0}^1 f(S_{t+1} \mid S_t = j) f(S_t = j \mid \psi_t)}, \end{aligned} \quad (41)$$

and then generate  $S_t$  using a uniform distribution. For example, we generate a random number from a uniform distribution between 0 and 1; if this number is less than or equal to the calculated value of  $\mathbb{P}(S_t = 1 \mid S_{t+1}, \psi_t)$ , we set  $S_t = 1$ , otherwise,  $S_t$  is set equal to 0.

In view of applications, let us now consider the following *four state MSM*:

$$\begin{aligned} y_t &\sim \mathcal{N}(0, \sigma_{S_t}^2), \quad t = 1, 2, \dots, T, \\ \sigma_{S_t} &= \sigma_1 S_{1,t} + \sigma_2 S_{2,t} + \sigma_3 S_{3,t} + \sigma_4 S_{4,t}, \\ S_t &\in \{1, 2, 3, 4\} \text{ with transition probabilities} \end{aligned} \quad (42)$$

$$p_{ij} := \mathbb{P}(S_t = j \mid S_{t-1} = i), \quad i, j = 1, 2, 3, 4,$$

where  $S_{k,t} = 1$  if  $S_t = k$ , otherwise  $S_{k,t} = 0$ . Note that this is a particular case of the model analysed in Section 2.1, where

$\mu_t = 0 \forall t$ , hence we can perform the procedure referred to serially uncorrelated data taking  $\mu_{S_t} = \mu = 0$  to start the Gibbs sampling algorithm, therefore we have

*Step 1.* Generate  $\tilde{S}_T := (S_1, S_2, \dots, S_T)$  conditional on  $\tilde{\sigma}^2 := (\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2)$ ,

$$\begin{aligned} \tilde{p} &:= (p_{11}, p_{12}, p_{13}, p_{21}, p_{22}, p_{23}, p_{31}, p_{32}, p_{33}, p_{41}, p_{42}, p_{43}), \\ \psi_T &:= (y_1, y_2, \dots, y_T). \end{aligned} \quad (43)$$

For this purpose, we employ the Multi-move Gibbs sampling algorithm:

- (1) run procedure in Section 2.1 with  $\mu_{S_t} = 0$  in order to get, from last iteration,  $f(S_T | \psi_T) = \mathbb{P}(S_T | \psi_T)$ ,
- (2) recalling that  $f(S_t | S_{t+1}, \psi_t) \propto f(S_{t+1} | S_t) f(S_t | \psi_t)$  for  $t = T-1, \dots, 1$ , we can generate  $S_t$  from the vector of probabilities

$$\begin{aligned} &(\mathbb{P}(S_t = 1 | S_{t+1}, \psi_t), \mathbb{P}(S_t = 2 | S_{t+1}, \psi_t), \\ &\mathbb{P}(S_t = 3 | S_{t+1}, \psi_t), \mathbb{P}(S_t = 4 | S_{t+1}, \psi_t)), \end{aligned} \quad (44)$$

$$Y_t^1 := \frac{y_t}{\sqrt{(1 + S_{2,t}h_2)(1 + S_{3,t}h_2)(1 + S_{3,t}h_3)(1 + S_{4,t}h_2)(1 + S_{4,t}h_3)(1 + S_{4,t}h_4)}} \quad (47)$$

and take  $Y_{1t} \sim \mathcal{N}(0, \sigma_1^2)$ , in (42). By choosing an *inverted Gamma prior distribution*, that is,  $f(\sigma_1^2 | h_2, h_3, h_4) \sim \text{IG}(\nu_1/2, \delta_1/2)$  where  $\nu_1, \delta_1$  are the known prior hyperparameters, it can be shown that the conditional posterior distribution from which we generate  $\sigma_1^2$  is given by:

$$\sigma_1^2 | \psi_T, \tilde{S}_T, h_2, h_3, h_4 \sim \text{IG}\left(\frac{\nu_1 + T}{2}, \frac{\delta_1 + \sum_{t=1}^T Y_t^1}{2}\right). \quad (48)$$

*Generating  $\bar{h}_2$  Conditional on  $\sigma_1^2, h_3$  and  $h_4$ .* Note that the likelihood function of  $h_2$  depends only on the values of  $y_t$  for which  $S_t \in \{2, 3, 4\}$ . Therefore, take  $y_t^{(1)} := \{y_t | S_t \in \{2, 3, 4\}, t = 1, \dots, T\}$  and denote with  $T_2$  the size of this sample. Then define

$$Y_t^2 := \frac{y_t^{(1)}}{\sqrt{\sigma_1^2 (1 + S_{3,t}h_3)(1 + S_{4,t}h_3)(1 + S_{4,t}h_4)}}, \quad (49)$$

hence, for the observation in which  $S_t = 2, 3$  or  $4$ , we have  $Y_{2t} \sim \mathcal{N}(0, \bar{h}_2)$ . If we choose an inverted Gamma distribution with parameters  $\nu_2, \delta_2$  for the prior, we obtain  $\bar{h}_2 = 1 + h_2$  from the following posterior distribution:

$$\bar{h}_2 | \psi_T, \tilde{S}_T, \sigma_1^2, h_3, h_4 \sim \text{IG}\left(\frac{\nu_2 + T_2}{2}, \frac{\delta_2 + \sum_{t=1}^{T_2} Y_t^2}{2}\right). \quad (50)$$

where, for  $i = 1, \dots, 4$ ,

$$\mathbb{P}(S_t = i | S_{t+1}, \psi_t) = \frac{f(S_{t+1}S_t = 1) f(S_t = i\psi_t)}{\sum_{j=1}^3 f(S_{t+1}S_t = j) f(S_t = j\psi_t)}. \quad (45)$$

*Step 2.* Generate  $\tilde{\sigma}^2$ , conditional on  $\tilde{S}_T$  and the data  $\psi_T$ .

We want to impose the constraint  $\sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2$ , so we redefine  $\sigma_{S_t}^2$  in this way:

$$\begin{aligned} \sigma_{S_t}^2 &= \sigma_1^2 (1 + S_{2,t}h_2) (1 + S_{3,t}h_2) (1 + S_{3,t}h_3) \\ &\times (1 + S_{4,t}h_2) (1 + S_{4,t}h_3) (1 + S_{4,t}h_4), \end{aligned} \quad (46)$$

where  $h_j > 0$  for  $j = 1, \dots, 4$ , so that  $\sigma_2^2 := \sigma_1^2(1 + h_2)$ ,  $\sigma_3^2 := \sigma_1^2(1 + h_2)(1 + h_3)$  and  $\sigma_4^2 := \sigma_1^2(1 + h_2)(1 + h_3)(1 + h_4)$ . With this specification, we first generate  $\sigma_1^2$ , then generate  $\bar{h}_2 = 1 + h_2$ ,  $\bar{h}_3 = 1 + h_3$  and  $\bar{h}_4 = 1 + h_4$  to obtain  $\sigma_2^2, \sigma_3^2$  and  $\sigma_4^2$  indirectly.

*Generating  $\sigma_1^2$ , Conditional on  $h_2, h_3$  and  $h_4$ .* Define for  $t = 1, \dots, T$

In case  $\bar{h}_2 > 1$  put  $h_2 = \bar{h}_2 - 1$  and  $\sigma_2^2 := \sigma_1^2(1 + h_2)$ . Otherwise reiterate this step.

*Generating  $\bar{h}_3$  Conditional on  $\sigma_1^2, h_2$  and  $h_4$ .* Operate in a similar way as above. In particular if we define  $y_t^{(2)} := \{y_t | S_t \in \{3, 4\}, t = 1, \dots, T\}$  we will obtain

$$Y_t^3 := \frac{y_t^{(2)}}{\sqrt{\sigma_1^2 (1 + S_{3,t}h_2)(1 + S_{4,t}h_2)(1 + S_{4,t}h_4)}} \sim \mathcal{N}(0, \bar{h}_3). \quad (51)$$

*Generating  $\bar{h}_4$  Conditional on  $\sigma_1^2, h_2$  and  $h_3$ .* Operate in a similar way as above. In particular if we define  $y_t^{(3)} := \{y_t | S_t = 4, t = 1, \dots, T\}$  we will have

$$Y_t^4 := \frac{y_t^{(3)}}{\sqrt{\sigma_1^2 (1 + S_{4,t}h_2)(1 + S_{4,t}h_3)}} \sim \mathcal{N}(0, \bar{h}_4). \quad (52)$$

*Step 3.* Generate  $\tilde{p}$  conditional on  $\tilde{S}_T$ . In order to generate the transition probabilities we exploit the properties of the *prior Beta distribution*. Let us first define

$$\begin{aligned} \bar{p}_{ii} &:= \mathbb{P}(S_t \neq i | S_{t-1} = i) = 1 - p_{ii}, \quad i = 1, 2, 3, 4; \\ \bar{p}_{ij} &:= \mathbb{P}(S_t = j | S_{t-1} = i, S_t \neq i), \quad i \neq j. \end{aligned} \quad (53)$$



Hence we have that

$$\begin{aligned} p_{ij} &:= \mathbb{P}(S_t = j \mid S_{t-1} = i) \\ &= \mathbb{P}(S_t = j \mid S_{t-1} = i, S_t \neq i) \mathbb{P}(S_t \neq i \mid S_{t-1} = i) \quad (54) \\ &= \bar{p}_{ij} (1 - p_{ii}) \quad \forall i \neq j. \end{aligned}$$

Given  $\tilde{S}_T$ , let  $n_{ij}$ ,  $i, j = 1, 2, 3, 4$  be the total number of transitions from state  $S_{t-1} = i$  to  $S_t = j$ ,  $t = 2, 3, \dots, T$  and  $\bar{n}_{ij}$  the number of transitions from state  $S_{t-1} = i$  to  $S_t \neq j$ .

Begin with the generation of probabilities  $p_{ii}$ ,  $i = 1, 2, 3, 4$  by taking the Beta distribution as conjugate prior: if we take  $p_{ii} \sim \text{Beta}(u_{ii}, \bar{u}_{ii})$ , where  $u_{ii}$  and  $\bar{u}_{ii}$  are the known hyperparameters of the priors, the posterior distribution of  $p_{ii}$  given  $\tilde{S}_T$  still belongs to the Beta family distributions, that is,

$$p_{ii} \mid \tilde{S}_T \sim \text{Beta}(u_{ii} + n_{ii}, \bar{u}_{ii} + \bar{n}_{ii}), \quad i = 1, 2, 3, 4. \quad (55)$$

The others parameters, that is,  $p_{ij}$  for  $i \neq j$  and  $j = 1, 2, 3$ , can be computed from the above equation  $p_{ij} = \bar{p}_{ij}(1 - p_{ii})$ , where  $\bar{p}_{ij}$  are generated from the following posterior Beta distribution:

$$\bar{p}_{ij} \mid \tilde{S}_T \sim \text{Beta}(u_{ij} + n_{ij}, \bar{u}_{ij} + \bar{n}_{ij}). \quad (56)$$

For example, given that  $p_{11}$  is generated, we can obtain  $p_{12}$  and  $p_{13}$  by considering

$$\begin{aligned} \bar{p}_{12} \mid \tilde{S}_T &\sim \text{Beta}(u_{12} + n_{12}, \bar{u}_{12} + \bar{n}_{12}), \\ \bar{p}_{13} \mid \tilde{S}_T &\sim \text{Beta}(u_{13} + n_{13}, \bar{u}_{13} + \bar{n}_{13}), \end{aligned} \quad (57)$$

where  $\bar{n}_{12} = n_{13} + n_{14}$  and  $\bar{n}_{13} = n_{12} + n_{14}$ . Finally, given  $p_{11}$ ,  $p_{12}$  and  $p_{13}$  generated in this way, we have  $p_{14} = 1 - p_{11} - p_{12} - p_{13}$ .

**Remark 2.** When we do not have any information about priors distribution we employ hyperparameters  $u_{ij} = 0.5$ ,  $i, j = 1, 2, 3, 4$ . Usually we know that elements of the matrix diagonal in the transition matrix are bigger than elements out of the diagonal, because in a financial framework regime switching happens only occasionally: in this case, since we want  $p_{ii}$  close to 1 and  $p_{ij}$ ,  $i \neq j$ , close to 0, we will choose  $u_{ii}$  bigger than  $\bar{u}_{ii}$ .

#### 4. Goodness of Fit

Since financial time series are characterized by complex and rather unpredictable behavior, it is difficult to find, if there is any, a possible pattern. A typical set of techniques which allow to *measure* the goodness of forecasts obtained by using a certain model, is given by the *residual analysis*. Let us suppose that we are provided with a time series of return observations  $\{y_t\}_{t=1, \dots, T}$ ,  $T \in \mathbb{N}^+$ , for which we choose, for example, the model described in (4) with  $M = 4$ . By running the procedure of Section 2.1 we obtain the filtered probabilities

$$\mathbb{P}(S_t = j \mid \psi_t), \quad j = 1, 2, 3, 4, \quad t = 1, \dots, T, \quad (58)$$

and, by maximization of the log-likelihood function, we compute the parameters  $\hat{\mu}_1, \dots, \hat{\mu}_4, \hat{\sigma}_1, \dots, \hat{\sigma}_4$ , therefore we

can estimate both the mean and variance of the process at time  $t$ , for any  $t = 1, \dots, T$ , given the information set  $\psi_t$  as weighted average of four values:

$$\begin{aligned} \hat{\mu}_t &:= \mathbb{E}(\mu_t \mid \psi_t) := \hat{\mu}_1 \mathbb{P}(S_t = 1 \mid \psi_t) \\ &\quad + \dots + \hat{\mu}_4 \mathbb{P}(S_t = 4 \mid \psi_t), \\ \hat{\sigma}_t^2 &:= \mathbb{E}(\sigma_t^2 \mid \psi_t) := \hat{\sigma}_1^2 \mathbb{P}(S_t = 1 \mid \psi_t) \\ &\quad + \dots + \hat{\sigma}_4^2 \mathbb{P}(S_t = 4 \mid \psi_t). \end{aligned} \quad (59)$$

If the chosen model fits well the data, then the standardized residuals will have the following form:

$$\hat{\varepsilon}_t = \frac{y_t - \hat{\mu}_t}{\hat{\sigma}_t} \sim \mathcal{N}(0, 1), \quad t = 1, \dots, T, \quad (60)$$

therefore it is natural to apply a normality test as, for example, the *Jarque-Bera* test (see [18]) for details. We recall briefly that Jarque-Bera statistics is defined as

$$JB := \frac{T}{6} \left( S^2 - \frac{1}{4} (K - 3)^2 \right), \quad (61)$$

where the parameters  $S$  and  $K$  indicate the skewness, respectively, the kurtosis of  $\hat{\varepsilon}_t$ . If  $\hat{\varepsilon}_t$  come from a Normal distribution, the Jarque-Bera statistics converges asymptotically to a chi-squared distribution with two degrees of freedom, and can be used to test the null hypothesis of normality: this is a joint hypothesis of the skewness being zero and the excess kurtosis  $(K - 3)$  being also zero.

**Remark 3.** Note that the Jarque-Bera test is very sensitive and often rejects the null hypothesis only because of a few abnormal observations, this is the reason why, one has to take point out these outliers which has to be canceled out before apply the test on the obtained smoothed data.

#### 5. Prediction

The forecasting task is the most difficult step in the whole MSM approach. Let us suppose that our time series ends at time  $T \in \mathbb{N}^+$ , without further observations, then we have to start the prediction with the following quantities:

- (i) the transition probability matrix  $P^* := \{p_{ij}\}_{i,j=1,2,3,4}$ ;
- (ii) the vector  $\pi_T := \mathbb{P}(S_T \mid \psi_T) = (\mathbb{P}(S_T = 1 \mid \psi_T), \dots, \mathbb{P}(S_T = 4 \mid \psi_T))$  obtained from the last iteration of the filter algorithm, for example, the procedure in Section 2.1.

It follows that we have to proceed with the first step of the filter procedure obtaining the one-step ahead probability of the state  $S_{T+1}$  given the sample of observations  $\psi_T$ , that is,

$$\mathbb{P}(S_{T+1} = j \mid \psi_T) = \sum_{i=1}^4 p_{ij} \mathbb{P}(S_T = i \mid \psi_T), \quad j = 1, 2, 3, 4. \quad (62)$$

Equation (62) can be seen as a prediction for the regime at time  $T + 1$ , knowing observations up to time  $T$ . At this

point, the best way to make prediction about the unobserved variable is the simulation of further observations. Indeed, with the new probability  $\mathbb{P}(S_{T+1} \mid \psi_T)$  and the vector of parameter estimates  $\hat{\theta} := (\hat{\mu}_1, \dots, \hat{\mu}_4, \hat{\sigma}_1, \dots, \hat{\sigma}_4)$  we can estimate the one step ahead mean and variance as follows:

$$\begin{aligned}\hat{\mu}_{T+1} &:= \mathbb{E}(\mu_{T+1} \mid \psi_T) \\ &:= \hat{\mu}_1 \mathbb{P}(S_{T+1} = 1 \mid \psi_T) + \dots + \hat{\mu}_4 \mathbb{P}(S_{T+1} = 4 \mid \psi_T), \\ \hat{\sigma}_{T+1}^2 &:= \mathbb{E}(\sigma_{T+1}^2 \mid \psi_T) \\ &:= \hat{\sigma}_1^2 \mathbb{P}(S_{T+1} = 1 \mid \psi_T) + \dots + \hat{\sigma}_4^2 \mathbb{P}(S_{T+1} = 4 \mid \psi_T).\end{aligned}\tag{63}$$

Then we simulate  $\hat{y}_{T+1}$  by the Gaussian distribution  $\mathcal{N}(\hat{\mu}_{T+1}, \hat{\sigma}_{T+1})$  and, once  $\hat{y}_{T+1}$  has been simulated, we define  $\psi_{T+1} := \{y_1, \dots, y_T, \hat{y}_{T+1}\}$ . Then we first apply again the filter procedure of Section 2.1 for  $t = T + 1$  in order to obtain  $\mathbb{P}(S_{T+1} \mid \psi_{T+1})$ , then we compute  $\mathbb{P}(S_{T+1} \mid \psi_{T+1})$ ,  $\hat{\mu}_{T+2}$  and  $\hat{\sigma}_{T+2}^2$ , and we simulate  $\hat{y}_{T+2}$  by the Gaussian distribution  $\mathcal{N}(\hat{\mu}_{T+2}, \hat{\sigma}_{T+2})$ . Latter procedure runs the same all the other time-steps  $T + 3, \dots, T + m$ , where  $m \in \mathbb{N}^+$  is the time horizon of our forecast.

*Remark 4.* We would like to underline that latter described method is not reliable with few simulations since each  $y_{T+\tau}$ , for  $\tau = 1, \dots, m$  may assume a wide range of values and a single drawn describes only one of the many possible paths. So we can think to reiterate previous strategy many times in order to compute the mean behavior of  $\mathbb{P}(S_{T+\tau} \mid \psi_{T+\tau})$ ,  $\hat{\mu}_{T+\tau}$  and  $\hat{\sigma}_{T+\tau}^2$ . After having obtained a satisfactory number of data, then we can construct a confidence interval within the state probability will more likely take value. Obviously a high number of iterations of latter procedure rapidly increases the computational complexity of the whole algorithm because of the MLE related computational complexity, therefore we will adopt a rather different strategy which consists in simulating  $y_{T+\tau}$   $N$  times at each step (e.g.,  $N = 10000$ ) and then taking the mean over those values. However, we must pay attention because the mean calculation could cancel the possible regime switching: for example, if we draw many times  $y_t$  from  $\mathcal{N}(0, \sigma_{S_t})$  and we take the mean, by the law of large number we will have zero at any time. To overcome this problem we can take the mean of absolute values and then multiply this mean by a number  $x$ , which is a random variable that takes values 1 or  $-1$ , with equal probability, hence deciding the sign of  $y_t$  at every simulation step.

## 6. Applications

In this section we are going to apply the classical inference approach for a MSM to analyse real financial time series. In particular we will first examine data coming from the Standard & Poor's 500 (S&P 500) equity index which is considered, being based on the 500 most important companies in the United States, as one of the best representations of the U.S. stock market. Secondly, we shall consider the DAX (Deutsche Aktien Index) index which follows the quotations of the 30 major companies in Germany. Our choice is motivated by

a twofold goal: first we want to test the proposed 4-states MSM model on two particularly significant indexes which have shown to incorporate abrupt changes and oscillations, secondly we aim at comparing the behaviour of the two indexes between each other.

Computations have been performed following the MSM approach described in previous section, namely exploiting the procedures illustrated in Section 2. Let us underline that, instead of a standard 3-states MSM model, we shall use a 4-states MSM approach both for the S&P 500 and the DAX returns. Moreover the analysis has been realized for different intervals of time, focusing mainly on the period of Global Financial Crisis.

**6.1. The S&P 500 Case.** Figure 1 reports the graph of the Standard & Poor's 500 from 1st June 1994 to 27th May 2014, and it points out the dramatic collapse of index prices in years 2008-2009, when the crisis blew-up causing the achievement, 6th of March 2009 with 683.38 points, of the lowest value since September 1996.

Because of the latter fact we decided to focus our analysis on recent years. In particular we take into account data starting from the 1st of June 2007, and until 27 May 2014, therefore, denoting with  $\Lambda$  the set of observations and with  $X_t$ ,  $t \in \Lambda$ , the price data of the S&P 500, returns are calculated as usual by  $y_t := (X_t - X_{t-1})/X_{t-1}$ ,  $t \in \Lambda$ , where  $\{y_t\}_{t \in \Lambda}$  are the values for which we want to choose the *best* MSM. Note that in our implementation we grouped the daily data into weekly parcels in order to make the filter procedures less time-consuming and have a more clear output, therefore we obtain a vector of 365 values, still denoted by  $y_t$ , as shown in Figure 2.

Next step consists in understand if the returns are serially correlated or serially uncorrelated, a task which can be accomplished by running some suitable test, for example, the Durbin-Watson test (see, for example, [19, 20] or [7]) computing directly the value of the autoregressive parameter  $\phi$  by least square methods, namely  $\phi := (\sum_{t \in \Lambda} y_t y_{t+1}) / (\sum_{t \in \Lambda} y_t^2)$ , which gives us a rather low value, that is,  $-0.0697$ , so that we can neglect the autoregressive pattern and start the analysis by considering S&P 500 returns to be generated by a Gaussian distribution with switching mean and variance, that is,

$$\begin{aligned}y_t &= \mu_{S_t} + \epsilon_t, \quad t \in \Lambda \\ \{\epsilon_t\}_{t \in \{1, \dots, T\}} &\text{ are i.i.d. } \mathcal{N}(0, \sigma_{S_t}^2), \\ \mu_{S_t} &= \mu_1 S_{1,t} + \dots + \mu_4 S_{4,t}, \\ \sigma_{S_t} &= \sigma_1 S_{1,t} + \dots + \sigma_4 S_{4,t}, \\ S_t &\in \{1, \dots, 4\} \text{ with probabilities } p_{ij},\end{aligned}\tag{64}$$

where for  $(k, t) \in \{1, \dots, 4\} \times \Lambda$  we have  $S_{k,t} = 1$  if  $S_t = k$ , otherwise  $S_{k,t} = 0$ . Therefore, we suppose that the state variable  $S_t$ ,  $t \in \Lambda$ , takes its values in the set  $\Omega := \{1, 2, 3, 4\}$ , and we expect that the probabilities of being in the third and fourth state increase as a financial crisis occurs. Exploiting the procedure provided in Section 2.1, with respect to the returns



FIGURE 1: Daily observations of S&amp;P 500 from 1994 to 2014.

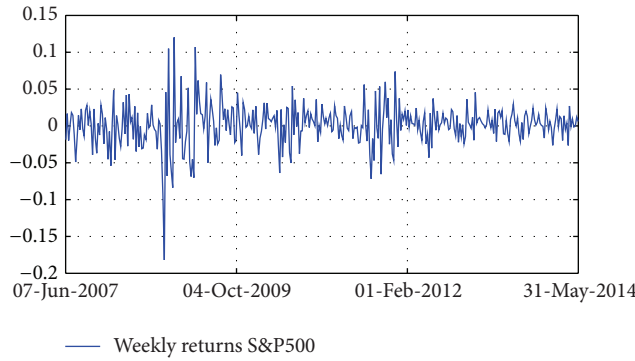


FIGURE 2: Daily returns of S&amp;P 500 from 2007 to 2014.

$y_t, t \in \Lambda = \{1, \dots, 365\}$ , we get the results shown in Figures 3 and 4.

Let us now consider the estimated standard deviation

$$\begin{aligned} \hat{\sigma}_t := \mathbb{E}(\sigma_t | \psi_t) &:= \hat{\sigma}_1 \mathbb{P}(S_t = 1 | \psi_t) \\ &+ \dots + \hat{\sigma}_4 \mathbb{P}(S_t = 4 | \psi_t), \quad t \in \Lambda, \end{aligned} \quad (65)$$

which we want to compare with the VIX index, also known as the *Chicago Board Options Exchange* (CBOE) market volatility index, namely one of the most relevant measure for the implied volatility of the S&P 500 index, whose value used by our analysis are reported in Figure 5.

What we obtain by plotting both estimated volatility and VIX values in the same graph can be seen in Figure 6, where the VIX trend is plotted in red, while we have used the blue color for the conditional standard deviation values.

Note that, in order to have values of the same order, each  $\hat{\sigma}_t, t \in \Lambda$ , has been multiplied by a scale factor equal to 1000. We would like to point out how the estimated standard deviation accurately approximates the VIX behaviour, hence allow us to count on an effective *substitute* for the volatility of the S&P 500, at least during a relative nonchaotic period. In fact we also underline that the greater discrepancies between real and simulated values appears during the maximum intensity period of the recent financial crisis. In particular the widest gaps are realized in correspondence with the recession experienced at the end of 2008.

In what follows we study how latter evidence influences the global goodness of the realized analysis. In particular we performed a *goodness of fit analysis* computing the standardized residuals of the proposed MSM by  $\hat{\epsilon}_t := (y_t - \hat{\mu}_t)/\hat{\sigma}_t, t \in \Lambda$ , where  $y_t$  is the observation of S&P 500 return at time  $t$ ,  $\hat{\mu}_t$  is the estimated conditional mean, and  $\hat{\sigma}_t$  is the standard deviation. If the model is a good fit for the S&P 500 return, standardized residuals will be generated by a standard Gaussian distribution. In Figures 7 and 8, we have reported both the histogram, its related graph, and the so called *normal probability plot* (NPP) for the standardized residuals.

Let us recall that the purpose of the NPP is to graphically assess whether the residuals could come from a normal distribution. Indeed, if such a hypothesis holds, then the *NPP has to be linear*, namely the large majority of the computed values, that is, the blue points in Figure 8, should stay close to a *particular* line, which is the red dotted one in Figure 8, which is the case in our analysis apart from the three points in the left-hand corner of the graph, which correspond to the minimal values of the vector of standardized residuals.

Applying two normality tests on  $\{\hat{\epsilon}_t\}_{t \in \Lambda}$ , that is, the *Jarque-Bera* test and (see, for example, [21, pag. 443]) the *Lilliefors* test, we have that the null hypothesis of normality for the standardized residuals can be rejected at the 5% level, unless the previous pointed out *outliers* are removed. Indeed if the two minimal standardized residuals, corresponding to  $\hat{\epsilon}_{71} = -3.8441$  and  $\hat{\epsilon}_{153} = -3.6469$ , are cancelled out from the vector  $\{\hat{\epsilon}_t\}_{t \in \Lambda}$ , previously cited tests indicate that the

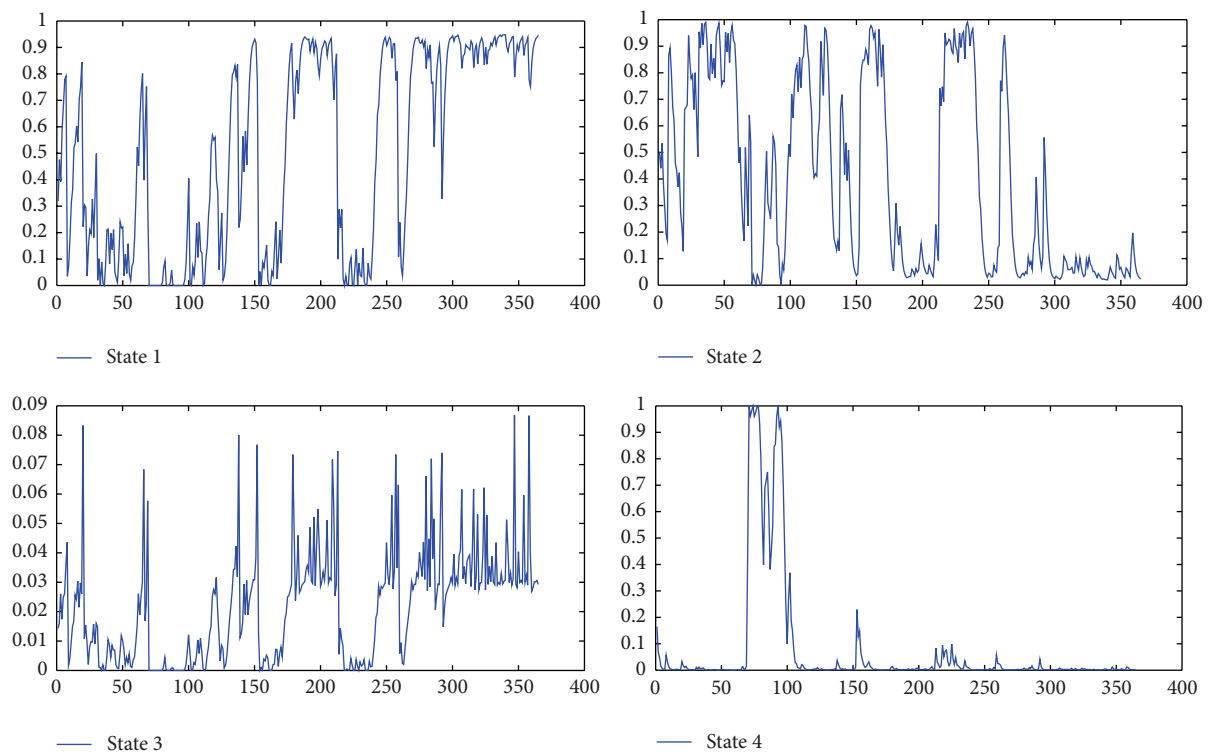


FIGURE 3: Filtered probabilities  $\mathbb{P}(S_t = k \mid \psi_t)$  for  $t \in \Lambda, k \in \{1, 2, 3, 4\}$ .

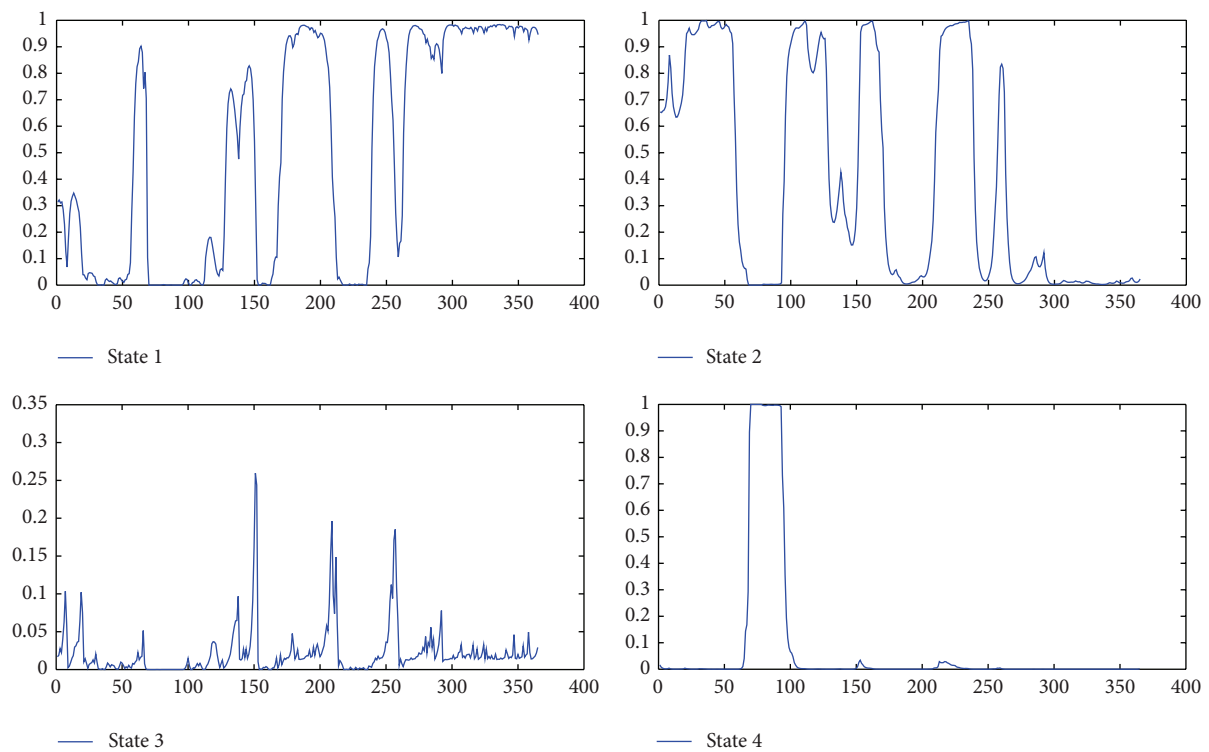


FIGURE 4: Smoothed probabilities  $\mathbb{P}(S_t = k \mid \psi_T)$  for  $t \in \Lambda, k \in \{1, 2, 3, 4\}$ .



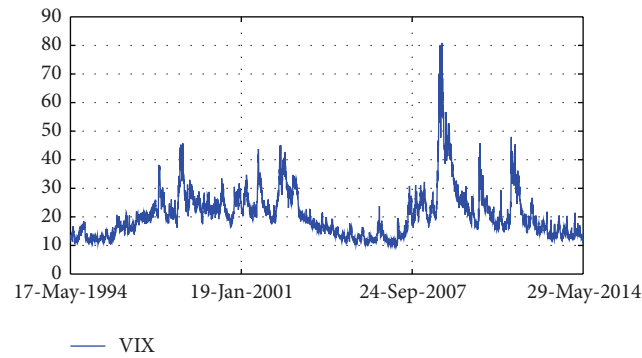


FIGURE 5: CBOE volatility index (VIX), daily data from 1994 to 2014.

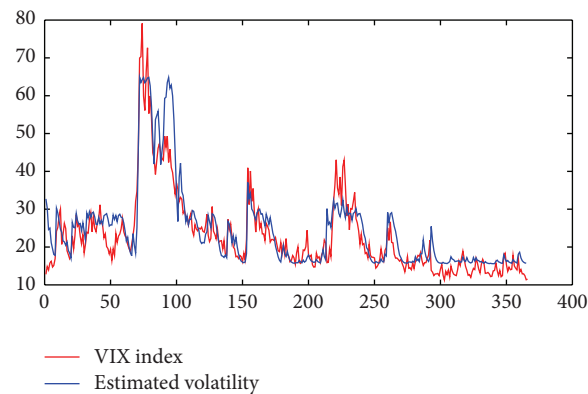


FIGURE 6: VIX index (red) versus estimated volatility (blue).

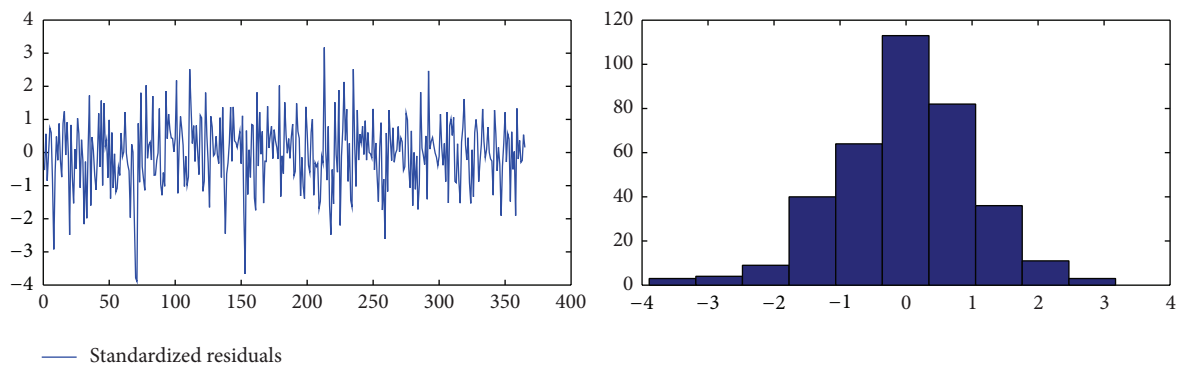


FIGURE 7: Plot and histogram of standardized residuals.

hypothesis of normality, at the same significance level of 5%, cannot be rejected. In particular, the Jarque-Bera statistics value is  $JB = 2.7858$  with corresponding  $P$ -value  $P_{JB} = 0.2153$ , and the critical value for this test, that is, the maximum value of the JB statistics for which the null hypothesis cannot be rejected at the chosen significance level, is equal to  $k_{JB} = 5.8085$ . Similarly, with regard to the Lilliefors test, numerical value of Liellifors statistics,  $P$ -value and critical value are respectively given by  $L = 0.0424$ ,  $P_L = 0.1181$  and  $k_L = 0.0472$ .

In what follows we develop the forecast procedure shown in Section 5. Since we are dealing with weekly data, let us

suppose we want to predict probability of volatility  $\hat{\sigma}_t$ ,  $t \in \Lambda$  on a time horizon of two months, hence 8 steps ahead, then simulations have been performed according to Remark 4, with  $N = 15000$ ,  $T = 365$ ,  $\tau = \{1, 2, \dots, 8\}$ , and  $x$  uniformly distributed in  $\{-1, 1\}$ . Obtained forecasting results are shown in Figure 9, where plots are referred to the observations from the 300th to the 373rd, with the last 8 simulated values within red rectangles.

**6.2. The DAX Case.** In what follows the proposed 4-state MSM shall be applied to analyse the *Deutsche Aktien Index* (DAX) stock index during a shorter, compared to the study

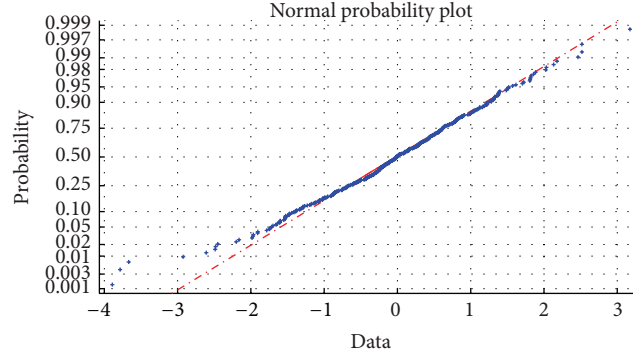


FIGURE 8: Normal probability plot of standardized residuals.

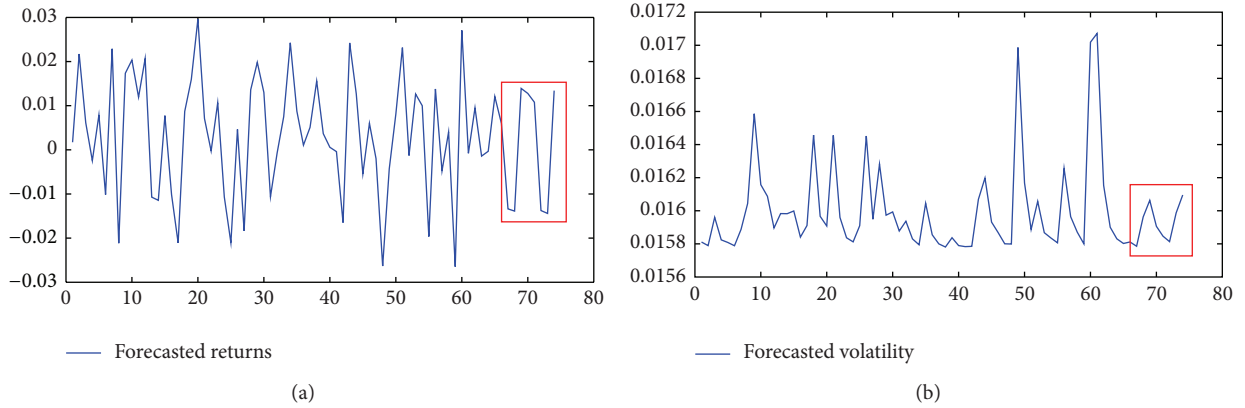


FIGURE 9: Forecast for returns (a) and conditional standard deviation (b).

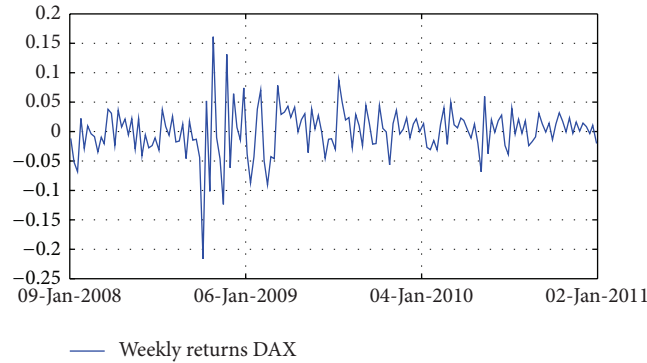
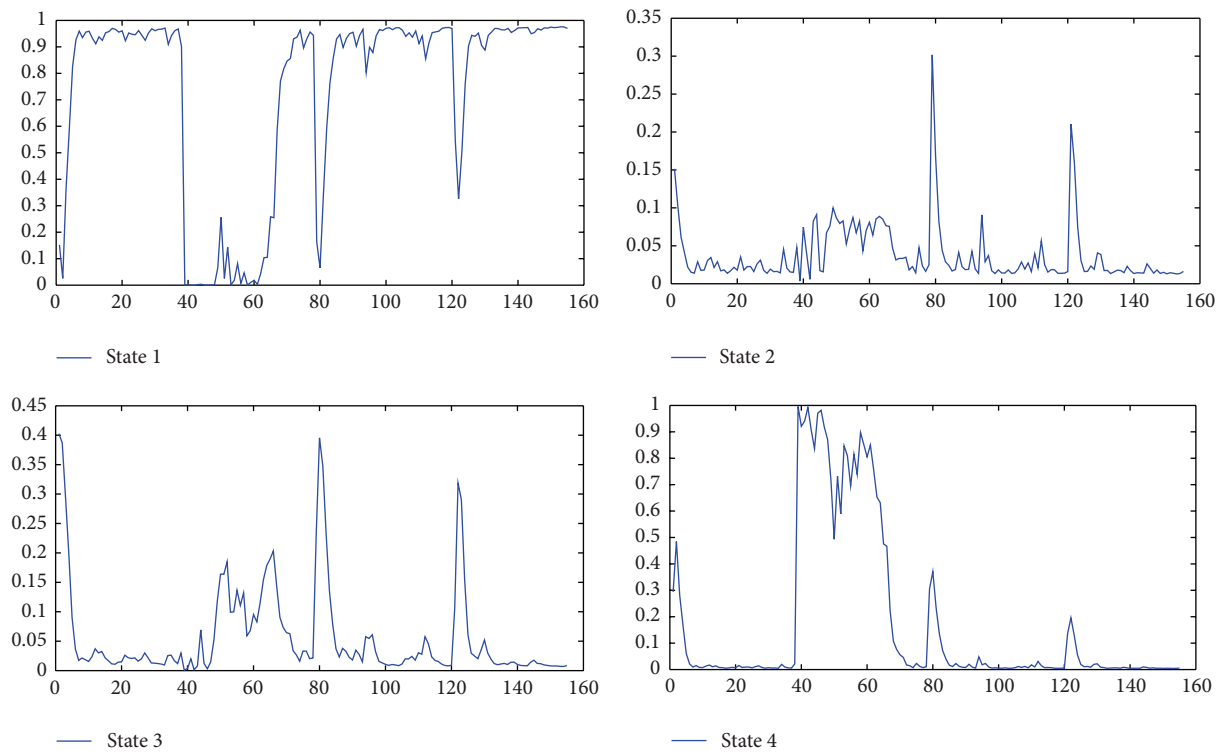
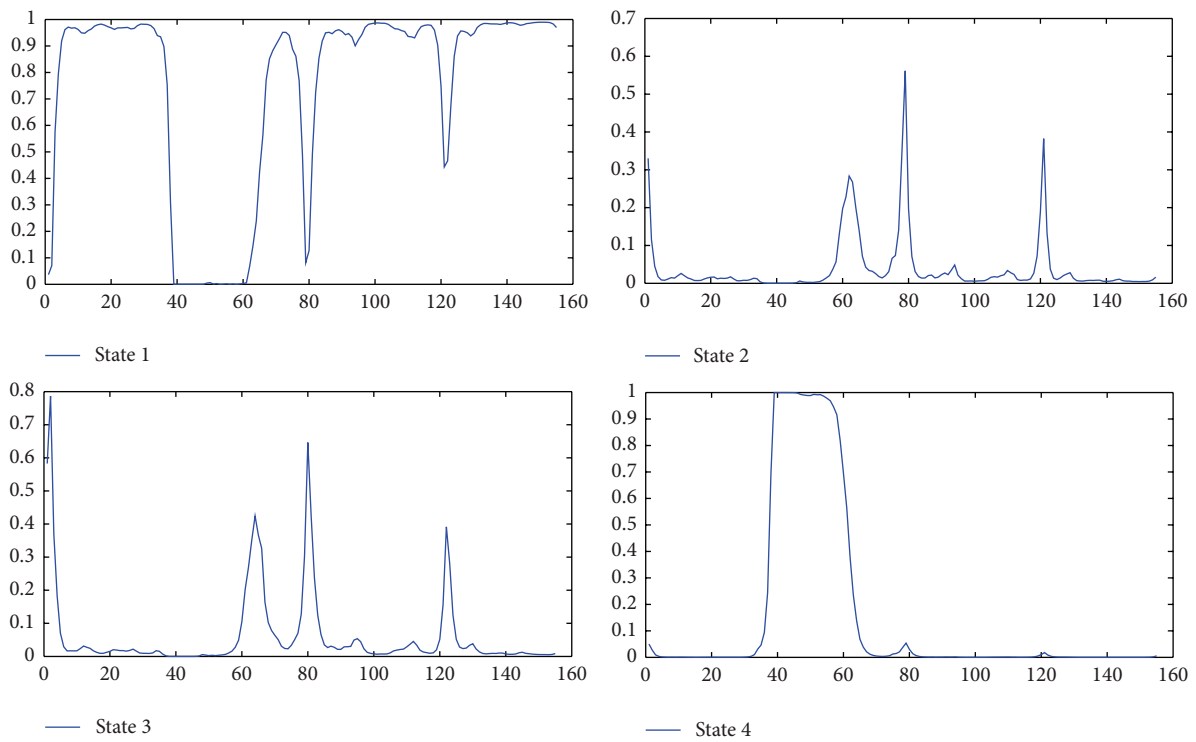


FIGURE 10: Weekly returns of DAX from 1st January 2008 to 1st January 2011.

made in Section 6.1, time interval, indeed we take into account data between the 1st of January 2008 and until the 1st of January 2011. Such a choice is motivated by the will of concentrate our attention on the most unstable period, that for the DAX index starts with the abrupt fall experienced by the European markets at the beginning of 2008. We underline that, with respect to the considered time period, the estimated autoregressive coefficient for the DAX index is a bit higher, in absolute value, than the one obtained for the S&P 500 index, indeed  $\hat{\phi}_{\text{DAX}} = -0.1477$  versus  $\hat{\phi}_{\text{S\&P500}} = -0.0697$ , so that we decided to fit the weekly DAX returns with an autoregressive

pattern using the procedure in Section 2.2 and according to the following MSM:

$$\begin{aligned}
 y_t &= \phi(y_{t-1}) + \epsilon_t, \quad t \in \Lambda \\
 \{\epsilon_t\}_{t \in \{1, \dots, T\}} &\text{ are i.i.d. } \mathcal{N}(0, \sigma_{S_t}^2), \\
 \sigma_{S_t} &= \sigma_1 S_{1,t} + \dots + \sigma_4 S_{4,t}, \\
 S_t &\in \{1, \dots, 4\} \text{ with probabilities } p_{ij},
 \end{aligned} \tag{66}$$

FIGURE 11: Filtered probabilities  $\mathbb{P}(S_t = k \mid \psi_t)$  for  $t \in \Lambda$ ,  $k \in \{1, 2, 3, 4\}$ .FIGURE 12: Smoothed probabilities  $\mathbb{P}(S_t = k \mid \psi_T)$  for  $t \in \Lambda$ ,  $k \in \{1, 2, 3, 4\}$ .

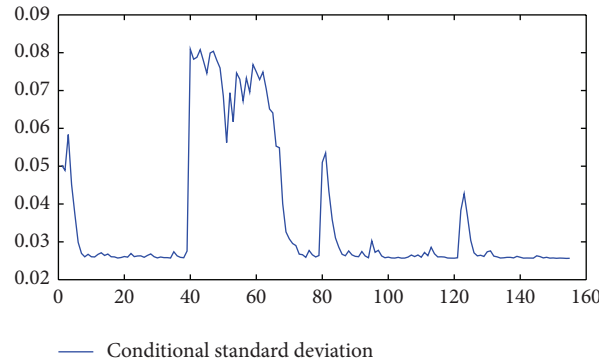


FIGURE 13: Estimated conditional standard deviation for DAX index.

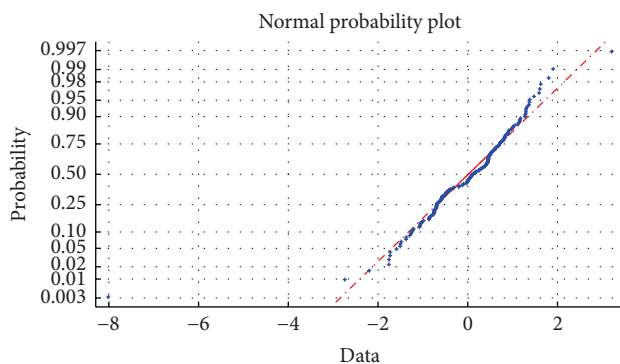


FIGURE 14: Normal probability plot for standardized residuals.

where  $\Lambda = \{1, 2, \dots, 156\}$  is the set of times,  $y_t$  is the return of DAX index at time  $t$  and for  $(k, t) \in \{1, \dots, 4\} \times \Lambda$  we have  $S_{k,t} = 1$  if  $S_t = k$ , otherwise  $S_{k,t} = 0$ . Note that we have slightly simplified the model keeping  $\mu_{S_t} = 0$ , that is we are dealing with a switching variance model where only  $\sigma_{S_t}$  switches in regime. The latter is not a restriction since we are interested in model the DAX index volatility. In Figure 10 we report the plot of weekly returns within the period of interest.

Figures 11, 12 and 13, show the graphical outputs of the performed analysis. In particular they respectively display both the filtered and smoothed probabilities for every regime, and estimated standard deviation  $\hat{\sigma}_t$ , for  $t \in \Lambda$ .

The *goodness of fit* analysis has been done applying normality tests to standardized residuals, with related NPP shown in Figure 14.

As in the S&P 500 case (see Section 6.1) both the Jarque-Bera and Lilliefors tests reject the null hypothesis of normality if they are performed taking into account the whole set of residuals, while, removing the single outlier corresponding to the smallest value of residuals (see the left-hand corner of the NPP in Figure 14) both tests are passed and we can conclude that our 4-state MSM provides a good fitting of examined data. In particular, as in the analysis of S&P 500, we computed numerical values for the two statistical test with a significance level  $\alpha = 5\%$ , obtaining  $JB = 0.2326$ ,  $P_{JB} = 0.500$ ,  $k_{JB} = 5.6006$  and  $L = 0.0712$ ,  $P_L = 0.0553$ ,  $k_L = 0.0720$ .

## 7. Conclusion

In this work we have shown how a four-state Regime Switching Model can be successfully exploited to study the volatility parameter which strongly characterizes concrete financial time series. In particular we have performed a deep analysis of data taken from the *Standard & Poor's 500* and the *Deutsche Aktien Index* series of returns focusing our attention on the most erratic periods for financial quantities influenced by the global financial crisis. We provided both accurate numerical results and related effective graphs, together with a detailed overview of the algorithms used to infer on parameters of the exploited switching models. Applications to the above mentioned markets (see Section 6) are completed by an accurate good of fitness study.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

- [1] H. Tong, *Non-Linear Time Series. A Dynamical System Approach*, Oxford University Press, Oxford, UK, 1990.
- [2] F. Bec, A. Rahbek, and N. Shephard, "The ACR model: a multivariate dynamic mixture autoregression," *Oxford Bulletin of Economics and Statistics*, vol. 70, no. 5, pp. 583–618, 2008.
- [3] J. Zhou and M. Chen, "Functional coefficient autoregressive conditional root model," *Journal of Systems Science & Complexity*, vol. 25, no. 5, pp. 998–1013, 2012.
- [4] S. M. Goldfeld and R. E. Quandt, "A Markov model for switching regressions," *Journal of Econometrics*, vol. 1, no. 1, pp. 3–15, 1973.
- [5] J. D. Hamilton, "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, vol. 57, no. 2, pp. 357–384, 1989.
- [6] C.-J. Kim and C. R. Nelson, *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, The MIT press, 1999.
- [7] J. D. Sargan and A. Bhargava, "Testing residuals from least squares regression for being generated by the Gaussian random walk," *Econometrica*, vol. 51, no. 1, pp. 153–174, 1983.



- [8] W. M. Bolstad, *Introduction to Bayesian Statistics*, Wiley-Interscience, Hoboken, NJ, USA, 2nd edition, 2007.
- [9] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, CRC Texts in Statistical Science Series, Chapman & Hall, London, UK, 2003.
- [10] V. Witkovský, "Computing the distribution of a linear combination of inverted gamma variables," *Kybernetika*, vol. 37, no. 1, pp. 79–90, 2001.
- [11] M. Chen, Q. Shao, and J. Ibrahim, *Monte Carlo Methods in Bayesian Computation*, Springer Series in Statistics, 2000.
- [12] P. Diaconis and S. Holmes, "Three examples of Monte-Carlo Markov chains: at the interface between statistical computing, computer science, and statistical mechanics," in *Discrete Probability and Algorithms*, vol. 72 of *The IMA Volumes in Mathematics and Its Applications*, pp. 43–56, Springer, New York, NY, USA, 1993.
- [13] D. Gamerman and H. F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Texts in Statistical Science, Chapman & Hall/CRC, New York, NY, USA, 2nd edition, 2006.
- [14] J. Albert and J. S. Chib, "Bayes inference via gibbs sampling of autoregressive time series subject to Markov mean and variance shifts," *Journal of Business & Economic Statistics*, vol. 11, pp. 1–15, 1993.
- [15] C. K. Carter and R. Kohn, "On Gibbs sampling for state space models," *Biometrika*, vol. 81, no. 3, pp. 541–553, 1994.
- [16] G. Casella and E. I. George, "Explaining the Gibbs sampler," *The American Statistician*, vol. 46, no. 3, pp. 167–174, 1992.
- [17] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [18] C. M. Jarque and A. K. Bera, "A test for normality of observations and regression residuals," *International Statistical Review*, vol. 55, no. 2, pp. 163–172, 1987.
- [19] J. Durbin and G. S. Watson, "Testing for serial correlation in least squares regression. I," *Biometrika*, vol. 37, no. 3-4, pp. 409–428, 1950.
- [20] J. Durbin and G. S. Watson, "Testing for serial correlation in least squares regression. II," *Biometrika*, vol. 38, no. 1-2, pp. 159–178, 1951.
- [21] W. J. Conover, *Practical Nonparametric Statistics*, John Wiley & Sons, 3rd edition, 1999.

